Experiments with time-frequency inversions

Jensen, Karl Kristoffer

Published in:
Frontiers of Research in Sound and Music

Publication date:
2009

Document Version
Early version, also known as pre-print

EXPERIMENTS WITH TIME-FREQUENCY INVERSIONS

Kristoffer Jensen

Department of Media Technology

Aalborg University Esbjerg, 6700 Esbjerg, Denmark

krist@aaue.dk

Keywords: Discrete Cosine Transform, Spectral and Temporal Envelope, Sonification, Kurtosis, Cepstrum.

Abstract.

All signals can become sounds. This is commonly called sonification (Kramer 1999). A particular method for obtained sounds is investigated here. Taken as the basis an existing sound, the transformed signal or shape of the signal is here considered as a sound itself. By calculating the spectral and temporal envelopes, and replacing one by the other, subtle but interesting effects are obtained. In contrast, by replacing the sound by the Discrete Cosine Transform of the same sound, a much more important change is produced. In particular, in the case of inharmonic or unvoiced sounds, such as the cymbal or the unvoiced consonants of speech, interesting textures are obtained. For most harmonic sounds, another harmonic sound is obtained by the DCT. If, finally, the envelope switching is done on the domain switched sound, some of the original qualities of the sound are re-introduced.

This work has been done for creative and pedagogical reasons. It extends the range of sound textures for contemporary music creation, and it also extends the number of means for understanding the implications of spectral and temporal envelopes, as well as the relationship between a signal and the Fourier/Cosine transform of the same signal.

1. Introduction

While time is sequentially perceived and frequency is parallel perceived, some similarity in the feature values of the two domains exists. This work proposes to investigate the utility and implications of conceptually switching domains. The domain switching can be done on the envelope level (switching the temporal envelope for the spectral envelope and vice verse), or directly on the samples or frequency bin values.

The temporal envelope (Hartmann 1997) is the shape of the temporal signal, and the spectral envelope is the shape of the spectrum. These envelopes often have similar decreasing shapes. In the case of envelope switching, if done on isolated musical sounds, it is changing the sound without altering the identity significantly. This is because both the temporal and the spectral envelope present a similar falling slope for many isolated sounds. In specific cases, such as the case of sustained sounds, e.g. the flute, the envelope inversion renders a more percussive sound. As for speech sounds, they are rendered in between, dependent on the actual speech pronounced (vowel/consonant, etc). The spectral/temporal envelope inversion can be considered as a filtering which modifies the sound without changing the identity of the sound.

In contrast to this, switching domain directly on the sample/bin level radically changes the sound. In order to simplify the domain switching, the discrete cosine transform is used. This retains the number of samples that are also real valued. When performing domain switching,

low pitches are transformed into high pitches and high pitches into low, but all kinds of changes occur, which renders the resulting sound unrecognisable.

Sonification is the task of using audio to convey information about non-audio signals (Kramer 1999). As such, it is used here to inform about the signification of the frequency domain signal of audio sounds. Sonification is often performed in order to facilitate the interpretation of the data analysed. Here, it is conceived to facilitate the interpretation of the signal, the transformed signal, and the process of transforming the signal.

This work has been done for two reasons, first to investigate alternative ways of better understanding the domain transforms of audio, and secondly as a means for creating sound using novel methods for use in creative works. Finally, this approach is also a means for presenting the details of both the components of the sound as well as the implication of the spectral domain in for instance pedagogical situations.

In section 2, the issues involved in estimating the spectral and temporal envelopes, performing the discrete cosine transform and normalizing the data are presented, while section 3 gives the details of the experiments performed with time/frequency inversions.

2. Technical issues

This section details the technical issues that exist within the time-frequency inversion experiments. These include the determination of the spectral and temporal envelopes, the domain switching algorithm, and the normalization involved in the domain switching that are performed in order to get similar kind of data in the different domains.

2.1. Spectral & temporal Envelope

A spectral (Smith 2009) or temporal (Hartmann 1997) envelope is created when the signal is said to have a faster evolving signal which is having the same strength throughout, and which strength is altered by a slowly varying signal, the envelope. Common methods for estimating the spectral envelope include the cepstrum alanysis (Bogert *et al* 1963), or the linear prediction (LPC) (Makhoul 1975). The faster evolving signal is the sound signal, considered static in the case of the temporal signal, or the peaks of the sinusoidal components in the spectral case. The envelope is related to the combined envelope of the individual component peaks in the case of the temporal signal, and the individual peaks of the sinusoidal components, in the case of the spectral envelope. It is clear that the distinction between the fast and slow evolving signal is a matter of definition. A standard real-time method for calculating the temporal and spectral envelope is by calculating the maximum of each time/frequency range. This range should be large enough so as to bridge the gaps between the (periodic) peaks of the temporal and spectral signals, while small enough so as to capture the important details of the envelopes.

In this work, however, the temporal and spectral envelope is calculated using the cepstrum (Bogert *et al* 1963) transform. This takes advantage of the fact that the Fourier transform represents slowly varying signals in the low coefficients and fast varying signals in the high coefficients. The spectral envelope is obtained by taking the inverse Fourier transform of the logarithm of the Fourier transform of the signal,

$$y = FFT^{-1}(|\log(FFT(snd))|). \quad (1)$$

This is known as the cepstrum alanysis (Bogert *et al* 1963). By setting the high order coefficients to zero, and inverting the steps,

$$se = e^{FFT^{-1}(y)}, \quad (2)$$

the spectral envelope is obtained. As the Fourier transformed signal is considered to have faster evolving sinusoidal peaks and a slower changing spectral envelope, the cepstrum low coefficients correspond to the spectral envelope, and the high coefficients correspond to the faster evolving components. By setting the high coefficients to zero, only the spectral envelope is retained. In a similar manner, the temporal envelope is obtained by taking the Fourier transform of the log of the sound directly,

$$y = FFT^{-1}(|\log(snd)|), \tag{3}$$

and zeroing out the high coefficients before doing the inverse transform. The cepstrum method yields approximative spectral and temporal envelopes, as for instance it doesn't respect the spectral peaks or the peaks in the temporal signal faithfully. It is still preferred over peak detection methods, as for instance, it gives less risks of falling into the in between peak holes.

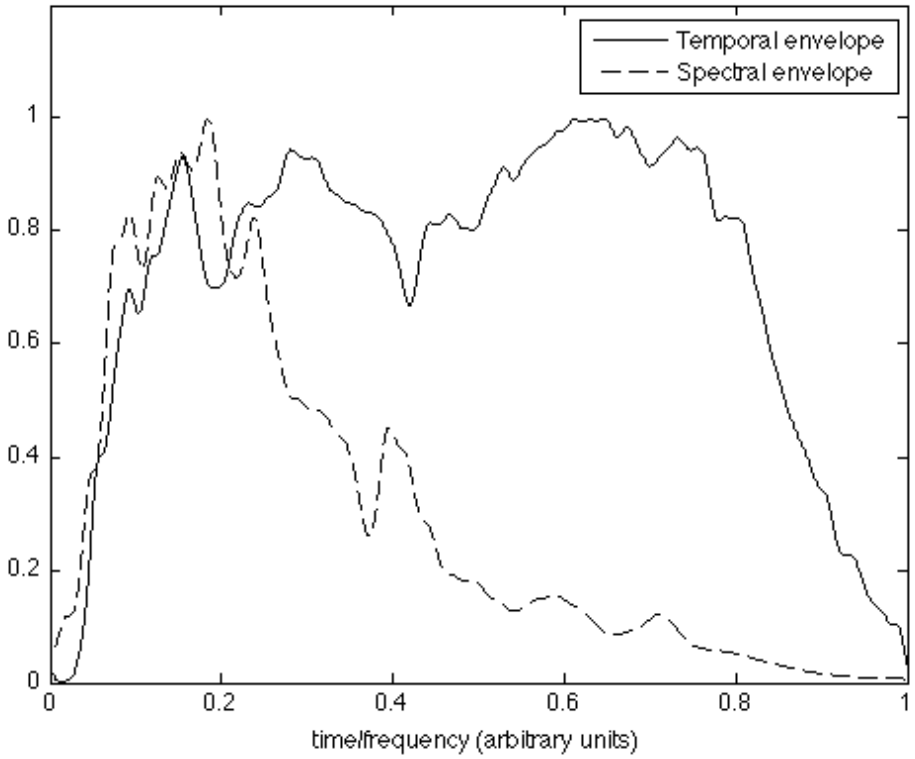

Figure 1. Temporal (solid) and spectral (stipled) envelopes for the trumpet sound.

As an example of the temporal and spectral envelope, those envelopes are shown in figure 1 for a trumpet sound. The x-axis is considered as both time and frequency with arbitrary units, as conceptually, the spectrum can be considered the time-domain signal and vice-versa. This trumpet sound is continuous, and therefore the spectral envelope is significantly weaker than the temporal envelope in the high time/frequency. This is true for continuous sounds, and music, but not for percussive sounds.

2.2. Discrete Cosine Transform

In order to obtain a real signal from what can be considered the spectral domain, the discrete cosine transform (DCT) is used. The DCT (Ahmed *et al* 1974) calculates a real signal closely related to the Fourier transform, but with real coefficients, and the same number of transform domain coefficients as the signal. As for the Fourier transform, low coefficients represent

slowly varying signals, and high coefficients represent fast varying signals. The DCT is defined as,

$$y(k) = w(k) \sum_{n=1..N} x(n) \cos \frac{\pi(2n-1)(k-1)}{2N}, \quad k = 1,...,N. \tag{4}$$

where

$$w(k) = \begin{cases} \sqrt{1/N}, & k = 1 \\ \sqrt{2/N}, & 2 \le k \le N \end{cases} \tag{5}$$

The DCT renders a dirac (pulse), if the original signal is periodic. As the DCT is a linear operation, each harmonic of a harmonic sound creates one pulse, which position $pos = 2 \cdot f \cdot L / sr$, where f is the frequency of the harmonic, and L is the length of the sound, in seconds, and where sr is the sample rate of the signal. The fundamental frequency of the dct signal is thus (when it is conceptually transformed back to the time domain),

$$f_0^{DCT} = sr / 2 \cdot f_0 \cdot L, \tag{6}$$

As the position of the pulses are proportional to the frequency of the partials, the amplitudes of the pulses are proportional of the amplitude, and the higher partials usually have weaker amplitudes, the resulting signal is generally percussive in nature.

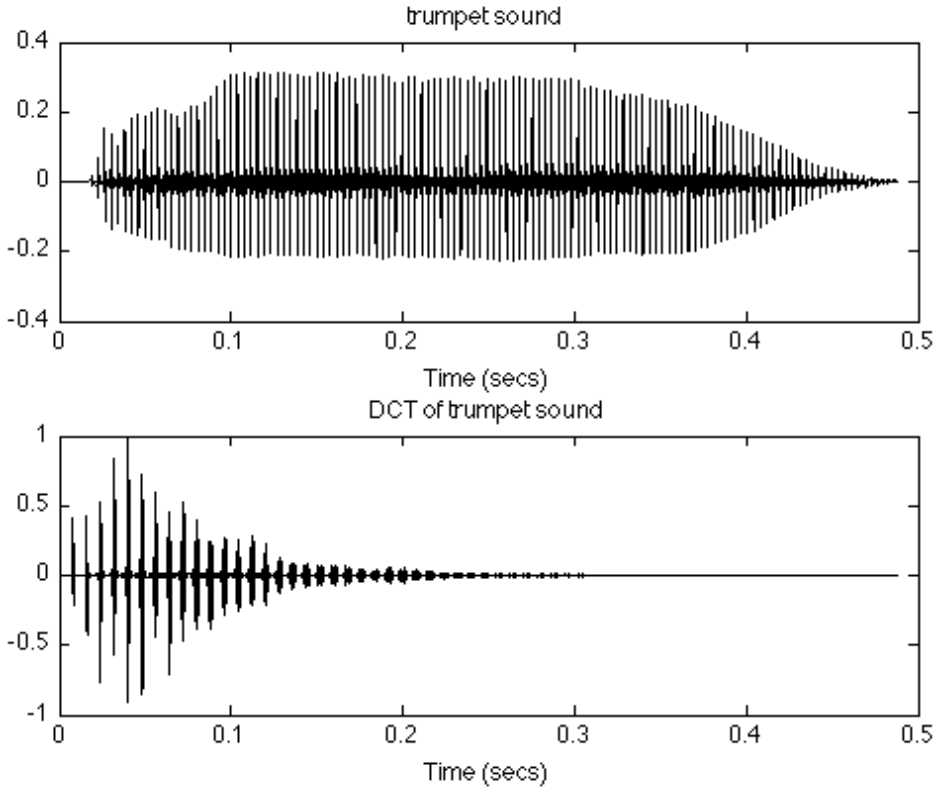

Figure 2. Trumpet sound (top), and DCT of trumpet sound (bottom).

In figure 2, the signal and the DCT of the signal of a trumpet sound is shown. As conceptually, the DCT is considered a sound, the x-axis is labelled as time in both cases. This is different from the cepstrum alanysis, in which the sound is transformed twice, thus returning to the time domain. The fundamental of the original sound is 262Hz, and that of the

DCT of the sound 120 Hz. The sample rate is 32000, and the length of the sound is approximately ½ seconds.

If the original signal is percussive, the DCT contains a sinusoid. By the linear nature of the DCT, a sum of pulses (which is a point-of-view of much noise) would generate a sum of sinusoids in the DCT domain. However, as Jensen (2004) demonstrates, both random pulses and the sum of sinusoids with random frequencies approach noise, as the number of pulses/sinusoids increases. Therefore, the noise signals, such as a cymbal, or whispering noise, does not render a voiced dct signal, generally.

2.3. Domain normalizations

In this section, the signals in the time and frequency domain are investigated, and the issues that need to be normalized in order to obtain an adequately sounding signal are identified and methods to normalize the signal are detailed. This involves the distribution of the signal. Indeed, if the sound is supposed to be Gaussian with zero skewness and kurtosis≈0, then the transform domain is usually also Gaussian with zero skewness, but a much higher kurtosis. A Gaussian signal has most values around the mean and gradually less values further away from the mean. The relative amount of the values close to the mean is called the standard deviation. If the peak is more or less peaked than an ordinary Gaussian signal, then the kurtosis is different from 0. While the skewness is not very much affected by the frequency domain transform, the kurtosis is very much so. Indeed, for the test signals investigated, except the Dirac signal, the transform domain kurtosis (mean kurtosis=2039) is much higher than the sound kurtosis (mean=11). The mean of the signal is not important and not seriously affected by the transform, and the standard deviation is identical in both domains (sound and dct of sound).

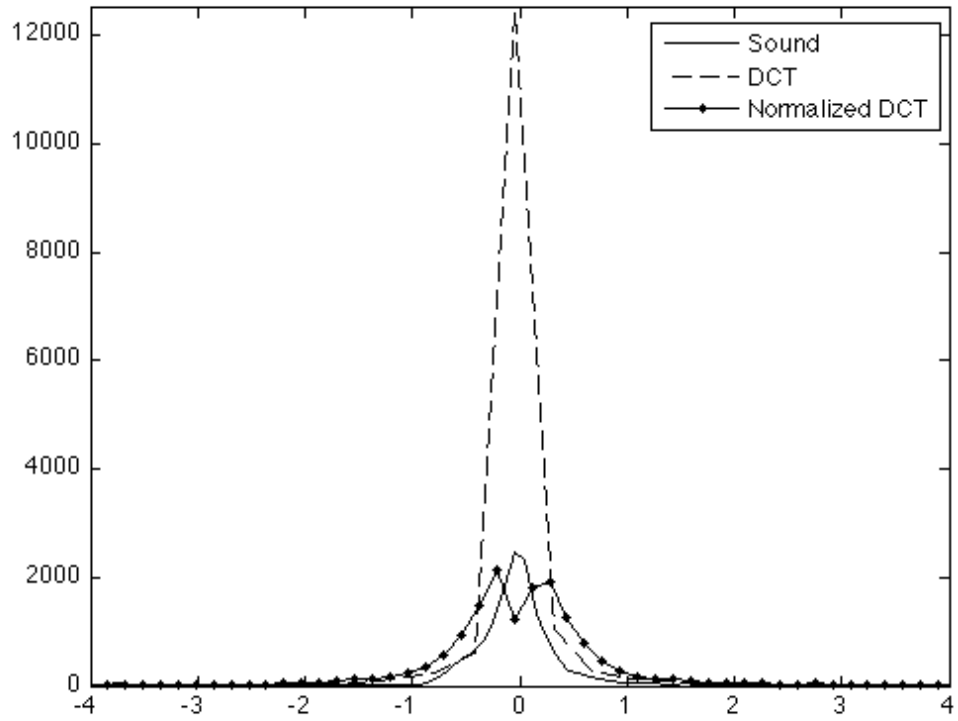

Figure 3. Distributions of sound, dct and kurtosis normalized dct of a trumpet sound.

A simple way to alter the kurtosis of a signal is to raise it to a power,

$$y = \mathrm{sgn}(x) \cdot |x|^p . \quad (7)$$

This effectively increases the difference between the low and high values, for $p>1$, and the opposite for $p<1$. In order to only modify the kurtosis, the standard deviation must be reset after eq. (7). This transformation effectively affects the small values, making them less probably if $p<<1$. This creates a platykurtic signal. The distribution of the original sound, the dct of the sound, and the kurtosis normalized dct (for a trumpet sound) is shown in figure 3. As can be seen, the lower values are less probably in the normalized dct.

An example of the kurtosis normalization ($p=0.48$) for the trumpet sound is shown in figure 4. The original kurtosis is 88, and after normalization it is 13.8.

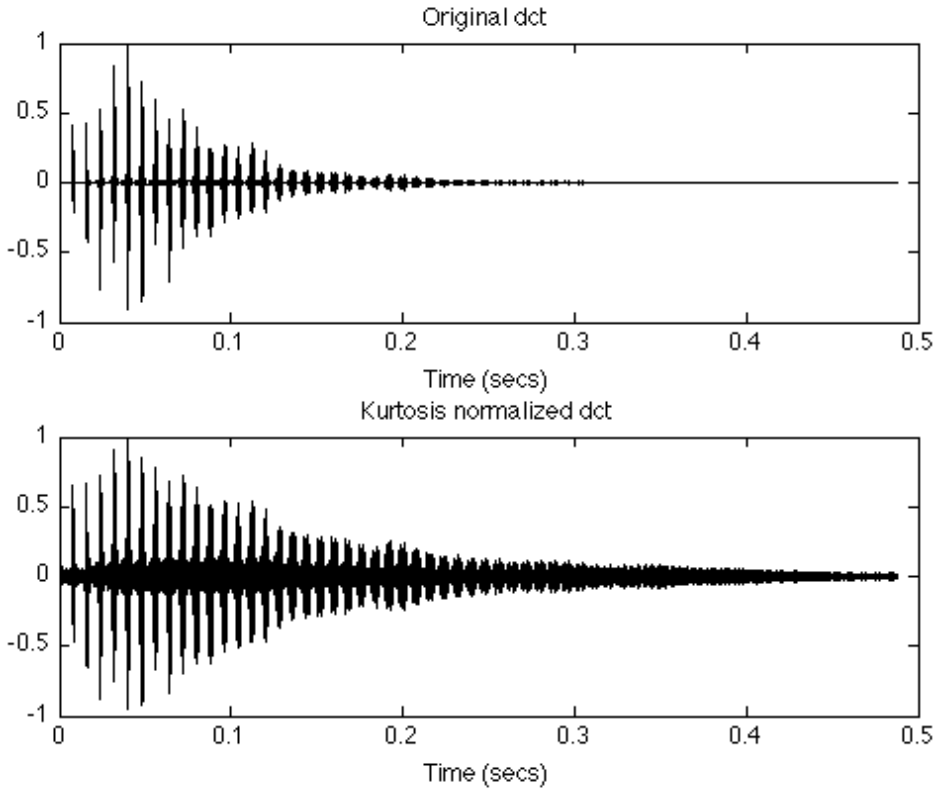

Figure 4. DCT and kurtosis normalized DCT of trumpet sound.

p is found experimentally in order to set the kurtosis of the transform domain (x) to that of the sound. In most cases, p is found to be between *0.1* and *0.4*. Effectively, this is a compression of the signal, and it often makes unvoiced noises emerge and drown voiced components. In the case of no content in the signal, the kurtosis normalization may strengthen quantization noise. For this reason, it is to be used with care. The eq. (7) can also be seen as a waveshaping (le Brun 1979) of the signal.

3. Experiments

In the experiments, a database of signals has been collected in order to submit them to different time/frequency inversion transformations. The database consists of isolated musical instruments from different classes, percussive, sustained, harmonic, stretched harmonic and inharmonic sounds. A number of speech signals are also included in the database. Finally, examples of complex music are also used in the experiments. By investigating the resulting

sounds from the different time/frequency inversions and related modifications, a better understanding of the possibilities these modifications have in terms of understanding of the transforms and in terms of use in creative works is obtained.

3.1. Envelope switching

In the envelope switching experiment, the temporal signal is giving the shape of the spectral envelope and the frequency domain signal is given the shape of the temporal envelope. Both these cases are obtained by simply multiplying the signal with the opposite domain envelope, and dividing by the same domain envelope. As a means for varying the signification of these transforms, the resulting signal is listened to.

As for isolated musical instrument sounds, both domains have similar shapes; the resulting sound is similar, with a slightly changed time evolution and also a modified spectral content. Generally, the sound is recognisable, with a change that is sometimes quite strange, but most often more subtle than strange. The most interesting isolated sounds are inharmonics sounds, like cymbals or bells, as in those sounds the spectral envelope is jagger, which renders a more *alive* sound, when used as the temporal envelope. As an example, the spectrogram of a cymbal sound is shown in figure 5.

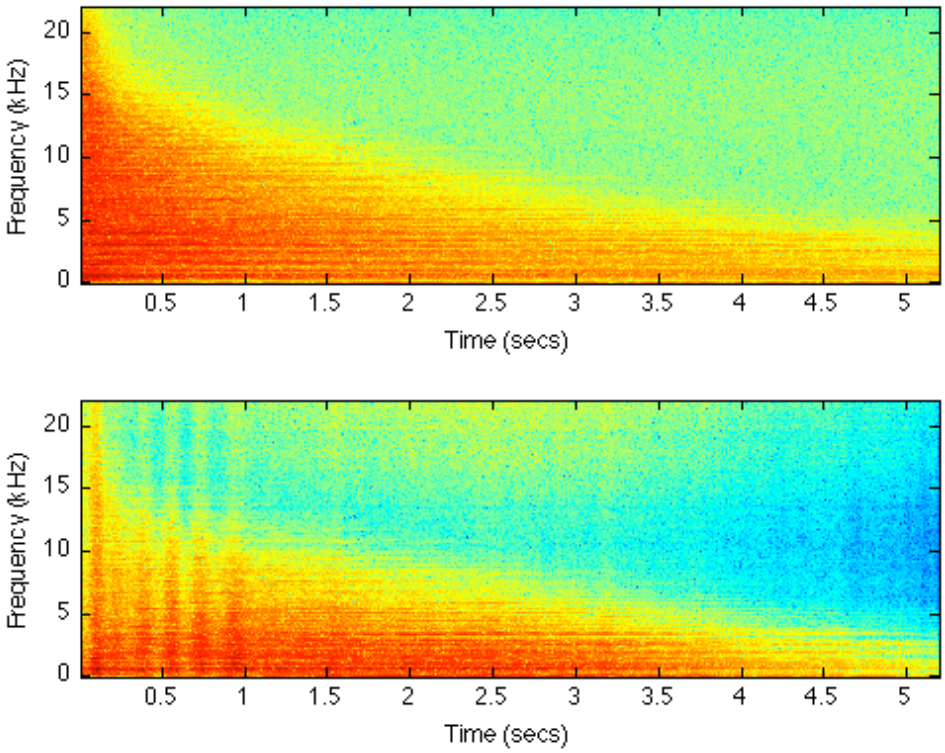

Figure 5. Spectrogram of cymbal sound. Original (top), and envelope domain inverted (bottom).

The original sound spectrogram is shown above in figure 5, and the modified sound, in which the temporal and spectral envelopes are swapped, is shown below in figure 5. Even if the sound retains approximately the same spectral and temporal range, only becoming perhaps marginally more percussive, it is rather changed with respect to the texture of the sound. Indeed, from a smooth voiced, albeit inharmonic, sound, the sound becomes fluctuating and not very voiced. Most instrumental sounds does not change that much, and most experimental sounds (sinusoids, pure noises, pulses) does not change at all.

3.2. Domain switching

The domain switching experiment renders quite different sounds as compared to the original sounds. In this experiment, the sound has indeed different texture and envelopes when obtained by taking the dct of the original sound. In this case, the high-pitched sounds becomes low-pitched, and vice-verse. As for the length of the original sound, if it is made longer, for instance with zeros, the resulting transform-domain sound has a lower resulting pitch. The pitched sounds always render a periodic sound while the inharmonic sounds, such as a cymbal renders noisy rich sounds. The speech sounds are rendered as strange sounds, with no recognisable identity, and a texture, which depends on the original sound, vowel or consonants. As explained above, the harmonic sounds render periodic sounds with pulses, all in all a quite bright sound. However, in many cases, the pulse periodicity descends below the pitch threshold, and become effectively individual pulses. This occurs when, assuming sr=44100, $f_0 > 1000/L$, approximately.

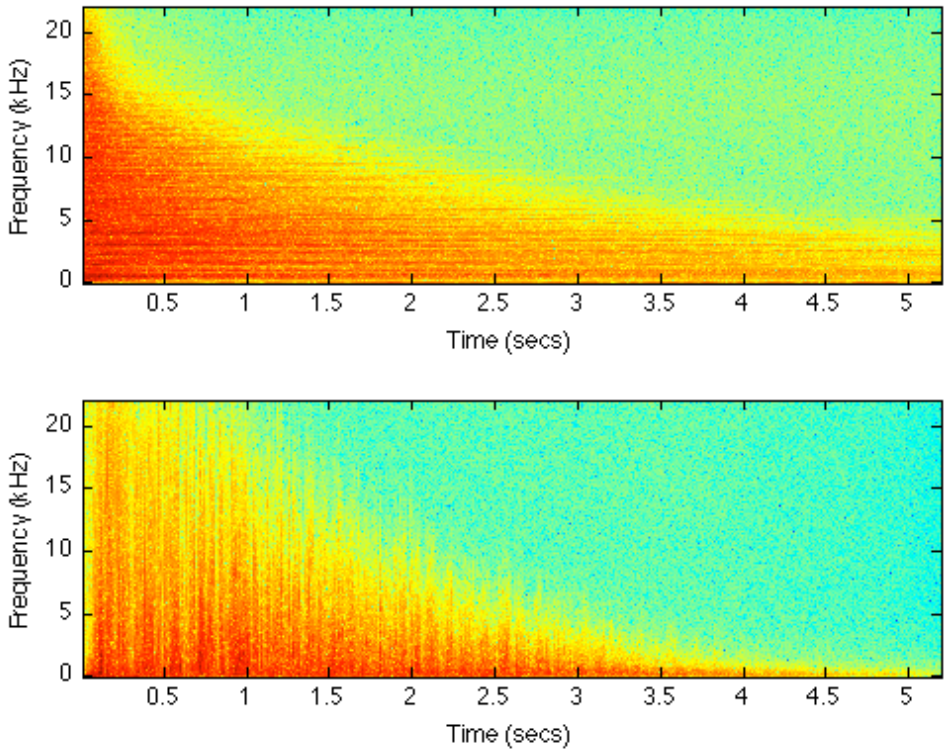

Figure 6. Spectrogram of cymbal sound. Original (top), and domain inverted (bottom).

The same cymbal sound as shown in figure 5 is shown in figure 6 but here the spectrograms depict the original and dct sounds. The dct sound is shown below the original sound. In this case, again the sound has retained the spectral and temporal range, becoming perhaps even more percussive, but the texture is totally changed. There is no trace of the original voiced components, and the uneven frequencies of the cymbal partials do not permit to detect any periodicity in the dct sound. It now has a texture resembling that of thunder.

3.1. Domain and envelope switching

The domain switched sound is also subjected to envelope switching. This implicates that the resulting sound now has a strange underlying signal (texture) with the original temporal and spectral envelopes. In this case, the effect of the modification of the envelope switching is larger than when it is done on the original sounds. The domain switched sound with original

spectral and temporal envelopes becomes more similar to the original sound. While the inverse domain sound often has a faster decay, i.e. the sound becomes more percussive, this effect is removed when the envelope switching is applied. Also, the inverse domain sound is generally less bright than the original sound, and the envelope switching also compensates this effect. Finally, some of the high-pitched sounds render very low-pitched transform domain sounds. These low periods are often confused as the temporal envelope, which considerably affects the resulting sound when the envelope switching is applied after the domain switching.

3.2. Kurtosis normalization

Many of the dct-modified sounds have very short pulses that are caused by the pure harmonics of the musical sounds. In addition, these sounds are often very percussive, because the original sounds have decreasing amplitude with frequency.

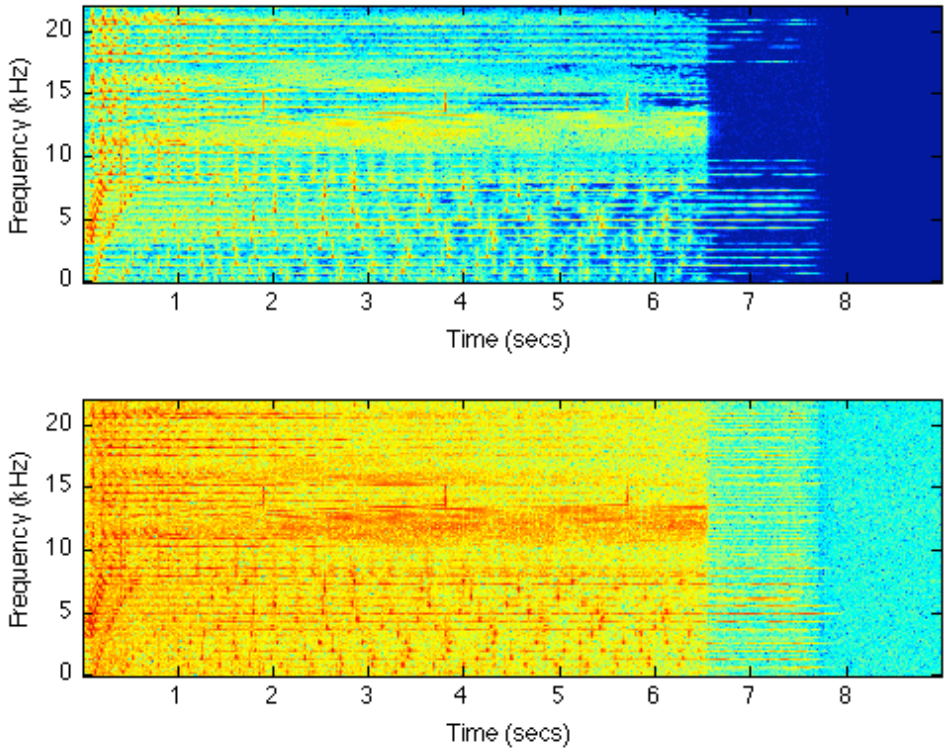

Figure 7. DCT of aq36. Before (top) and after (bottom) kurtosis normalization.

The results of this are percussive (short) sounds with pulses. If these sounds are subjected to the kurtosis normalization of eq. (7), more of the underlying texture is becoming perceptible. This can be noise caused by the blowing of the flute, or from the bowing of the string, or it can be sinusoids caused by pulses in the original sound. Often, unfortunately, there is only quantization noise, which is amplified by the kurtosis normalization. The kurtosis normalization seems most appropriate when applied to the dct of complex music. As an example, the spectrogram of the dct of 9 seconds of the Chinese pop song *aq36* (2005) is shown in figure 7.

It is clear that the upper original dct has more sparse components, and faster decreasing energy in the end. In comparison, the kurtosis normalized dct (bottom) has more energy throughout and generally is perceived as being more *dynamic*.

3.3. Convolutions and Multiplications

As convolutions and multiplications are two sides of the same effect (multiplication in the time domain corresponds to convolution in the frequency domain and vice-versa), these effects have also been incorporated into the experiment. The multiplication or convolution with the time domain or the dct signal has not, however interesting, been performed here. Instead, only the cross-domain convolution and multiplication is done. The convolution has as main effect a massive reverberate effect. As the dct signal often has many pulses, it effectively copies the time signal many times, which increases the reverberation. Generally, the time domain signal is better retained in the convolution. Most of the dct signal is lost after convolution, i.e. it is not perceptible anymore. The multiplication in comparison often retains the texture of the dct signal, which is rendered percussive by the time signal. This is particularly clear in the case of music, in which case, the rhythmic structure seems retained in the inharmonic and unvoiced dct textures. The same Chinese pop song as in figure 7 is shown in figure 8.

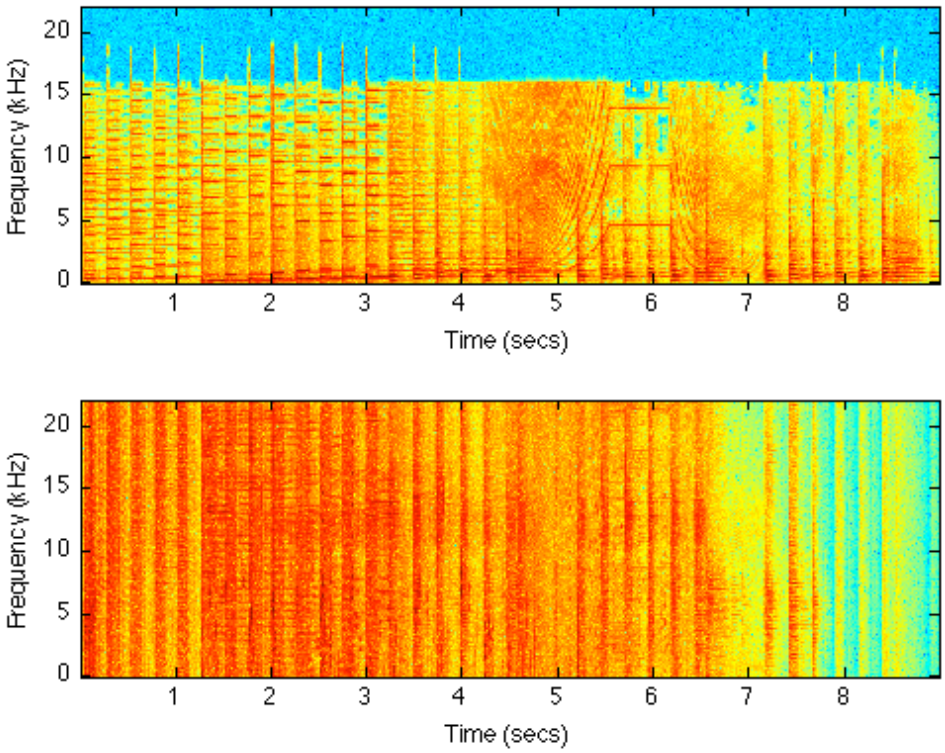

Figure 8. Spectrogram of 9 seconds excerpt of Aq36. Original (top) and multiplication of original and dct of original (bottom).

It is clear that the onset of the original signal is retained in the cross-domain multiplied signal. Some hints of the original toned material are also visible in figure 8. Perceptible, this is most clear in the end of the excerpt. The cross-domain multiplied signal has interesting characteristics that seem useful in creative endeavours in particular.

4. Conclusions

The Fourier transform and related algorithms (Cosine transform, etc…) are used in many multimedia applications. Two fundamental ways of understanding the time/frequency

transforms are by mathematics by transform pairs, or by seeing the transform as a filter bank, which improve the visibility of the sinusoids in the signal. In all practical implementation of transforms, the problem of time/frequency uncertainty is another important issue that needs to be taken into consideration.

Here, as a novel approach, the sonification of the Discrete Cosine Transform permits the understanding of key elements of the transform; sinusoids are transformed into peaks, noise into noise, and the transform is linear. Inversion of the temporal and spectral envelopes permits the understanding of the decreasing of the spectral envelope with frequency, and how this is related to the percussive nature of many musical sounds. In contrast to this, complex music is found to have continuous spectrum with more high frequency energy, and also a continuous temporal envelope. The perceptual analysis of the inversions of these envelopes (replacements) allows the discrimination between the envelopes and the texture of the sound.

In order to permit the time/frequency envelope inversions, a homogenous temporal and spectral envelope estimator based on the cepstrum alanysis and a novel time-domain equivalent, is introduced.

As a method for creation of musical sounds, this method is also very promising. The dct renders simple interesting sounds, if based on instrumental sounds, and rich, complex sounds, if based on complex music. The envelope switching creates more subtle changes in the sounds, while the kurtosis normalization method allows the creation of a range of sounds with varying dynamic range. A low kurtosis value renders rich sounds, while a high value renders more dynamic sounds with a sparse content.

In the long-term, the sonification of time/frequency transforms will be investigated and possibly integrated into the curriculum of signal processing topics for creative purposes. In the short-term, the work presented here is used as the basis for a music composition in collaboration with composer Laurent ‘Saxi’ Georges.

5. References

Ahmed, N., T. Natarajan, and K. R. Rao, *Discrete Cosine Transform*, IEEE Trans. Computers, pp 90-93, Jan 1974.

Bogert, B. P., M. J. R. Healy, J. W. Tukey. *The quefrency alanysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking*. in Rosenblatt, ed., Time Series Analysis. pp 209-243, New York: Wiley. 1963.

Hartmann, W. M., Signals, sound, and sensation. Springer, 1997

Jensen, K., *Atomic Noise*, Organised Sound, 10(1) pp 75-81. 2005.

Kramer. G., Nsf sonification report. March 1999. Available online: http://dev.icad.org/node/400, accessed 10/11 2009.

le Brun, M., *Digital waveshaping synthesis*, J. Audio Eng. Soc. 27(4), pp 757-768, April 1979.

Makhoul. J., *Linear prediction: A tutorial review*. Proceedings of the IEEE, 63(5). pp 561–580, 1975.

Smith, J. O. *Spectral Audio Signal Processing*, March 2009 Draft, http://ccrma.stanford.edu/~jos/sasp/, online book, accessed 10/11 2009